Available Admission Capacity estimations in IEEE 802.11 Access Points

Technical Report

Eduard Garcia
Wireless Networks Group, Entel Dept., UPC,
Avda. del Canal Olímpic, 15, 08860 Castelldefels, Spain
Phone: +34 93 413 7218; Fax: +34 93 413 7007

May 2008

ABSTRACT

This technical report is intended to provide extended information about the Available Admission Capacity estimation mechanism for IEEE 802.11 cells previously presented in [1]. The original formulation has been revised in this paper avoiding unnecessary approximations. Furthermore, a comprehensive evaluation of the algorithm is also provided.

I. INTRODUCTION

The present popularity of WLANs, especially those based on IEEE 802.11 standards, requires special attention with the aim of explaining all observable phenomena which affect their efficiency and their provision of quality. IEEE 802.11 standards define several sets of modulations and coding rates for the different physical layers. Each different scheme provides a different transmission rate, but the higher the chosen rate, the worse it performs in presence of noise and interference. As the quality of the signal gets worse, the physical rate must be lowered in order to achieve an acceptable packet error ratio (PER). Furthermore, a CSMA-type access is used: a station wishing to transmit first probes the medium and transmits only if the medium is sensed idle, ensuring long-term channel access fairness among all active stations. In a multi-rate environment, this concept of fairness involves a considerable loss of efficiency for the whole cell, since "slow" stations will capture the common medium for longer periods. In this way all stations obtain an equal bandwidth (expressed in Bytes/s). In consequence fairness is achieved at the cost of penalizing higher rate stations, which leads to a loss of efficiency.

In [1] we provided an analytical model that is revised and validated in this report through practical measurements and simulations. From the proposed formulation in [1], an algorithm was also derived, developed to run in commercial IEEE 802.11 Access Points. The algorithm performs an estimation of the maximum bandwidth that the Access Point (AP) can offer to each of its associated stations. The results of this mechanism can also be used to compute a load metric based on time share: the Available Admission Capacity (AAC), which describes the actual load of a cell according the capacity that an AP can offer to a new station. The use of our analytical

method to provide the capacity that a network is able to offer to a given station avoids the undesired effects of existing invasive mechanisms (e.g. see [2]), which consist of injecting real data over the air interface, thus consuming valuable resources and possibly producing a negative impact on ongoing real-time transmissions [3][4].

II. THROUGHPUT ESTIMATION FOR A NEW STATION

Straightforward measurements can be used to derive the throughput that is currently devoted to a station with ongoing data transmissions. However, more detailed study is required to determine how to predict the throughput that a new station will obtain before it actually starts to transmit or the maximum bandwidth that can be allocated to a station if it increases its offered traffic in a multi-rate cell. In a previous work [1], we proposed an algorithm that can accurately calculate the throughput available to a new station, which is obtained from AAC. This algorithm is based on the assumption that the IEEE 802.11 MAC maintains fairness in terms of access probability independently of the rate and bandwidth requirements of each station. It also takes into account the inherent "performance anomaly" [5] in multi-rate CSMA/CA networks. Based on these statements, we define T_{cycle} as the average time required to send one frame from each of the competing stations:

$$t_i = DIFS + T_{data}(i) + SIFS + T_{ACK} \quad (1)$$

$$T_{cycle} = \max(TBO_i) + \sum_{i=1}^{n} t_i \quad (2)$$

Note that t_i is defined for a basic CSMA/CA access; if RTS/CTS handshake is used, $2 \cdot SIFS$, T_{RTS} and T_{CTS} must be added to t_i. The duration of the data frame is T_{data}. TBO_i represents the average time during which a station i waits for the backoff timer to expire before attempting to transmit. Note that under saturation conditions, backoff slots are shared. Therefore, the idle time spent on backoff within a T_{cycle} is equivalent to the largest average backoff in the cell. Without saturation, stations are not "synchronized" and the probability that two stations share their backoff periods (or a part of them) is lower. Consequently, without saturation, our approximation introduces some error. In the next section we show that the error introduced is small.

TBO_i depends on the number of previous transmission attempts. The average value of the backoff interval after j consecutive transmissions is given by:

$$T_{BO}(j) = \begin{cases} \frac{2^j (CW_{\min} + 1) - 1}{2} T_{slot} & 0 \le j \le 6 \\ \frac{CW_{\max}}{2} T_{slot} & j \ge 6 \end{cases} \quad (3)$$

We consider that the number of packet retransmissions required to successfully transmit a single packet is a geometrically distributed random variable. If P_i is defined as the probability that a frame sent by i has to be retransmitted (including the effects of collisions and channel errors), the average backoff interval for station i is:

$$TBO_i = \sum_{j=0}^{\infty} (1 - P_i) P_i^j \cdot T_{BO}(j) \quad (4)$$

Note that from (4) we have to compute an infinite series. However, its operation can be stopped when the required precision is met. For example, for typical P_i values ranging from 0.1 to 0.2 and a required precision of 1 μs, 6 to 10 iterations are enough.
T_{data} can be further decomposed into[1]:

$$T_{data}(i) = T_{preamble} + \frac{8 \cdot (H_{MAc} + MSDU_i)}{r_i} \qquad (5)$$

where r_i is the physical bit rate at which node i sends data frames with a payload of $MSDU$ bytes. The remaining values vary according to the standard and the modulation used.

We define the overhead produced in layers 1 and 2 of a node i as $OH_i = r_i \cdot t_i / (8 \cdot MSDU_i)$. In order to calculate the actual TO (traffic offered) it is necessary to consider this overhead and the possible retransmissions given the traffic offered by the upper layers (TO_{app}):

$$TO_i = \frac{OH_i}{1 - P_i} TO_{app}(i) \qquad (6)$$

Resource distribution in WLAN provides a Max-min fairness in which small flows receive the volume they demand and larger flows share the remaining capacity equally. In other words, all stations whose traffic (TO_i) is equal to or smaller than the bandwidth that should be allocated under saturation conditions will be able to carry all of their offered traffic (TO_{app}). Saturated stations will share their corresponding bandwidth and the excess time that is not used by the non-saturated nodes. The algorithm is defined as follows:

Algorithm 1: AAC Algorithm

```
Tcycle'←Tcycle
OrderIncr(N, δi)
for all i ∈ N do
   if δi ≤ 1
      Li ← TOi/ri
      Texc ← (1-δi)/ ri
      Tcycle' ← Tcycle' – 1/ ri
      for all j ∈ N and δj > 1 do
         Lj ← Lj(1+Texc/ Tcycle')
         Sj ← Ljrj
         δj ← TOj/Sj
      end for
   end if
end for
```

where S_j is the throughput that station j would obtain in saturation conditions. The parameter δ_i is defined as the proportion of the maximum throughput that can be achieved by station i which is actually used: $\delta_i = TO_i/S_i$. Before the first execution, stations are ordered according to increasing value of the parameter δ. Stations with $\delta \leq 1$ use fewer resources (or an equal number) than those that would be allocated under saturation conditions, and their TO_{app} will therefore be carried. The proportion of time that is not used by non-saturated stations (T_{exc}) will be divided fairly between the remaining stations according to a new time T_{cycle}' in which the stations that have already been served are not considered. Note that greedy applications are modeled with $\delta_j > 1$ regardless of the

[1] case of DSSS-CCK modulations, for more details on OFDM see [6].

value of the S_j parameter. The *TO* of a new station is not known in advance and the maximum throughput it can potentially achieve is therefore calculated in saturation. The value of L_i represents the individual load contributed by node *i* defined as the proportion of T_{cycle} that is used by the *ith* node.

By using this formulation it is possible to perform the capacity estimation in real time, assuming that the required statistics are updated regularly by the firmware/driver of the wireless interface. Consequently, this estimation can handle varying traffic demands and varying channel conditions, since it also takes into account the effect of collisions and errors produced by noise and interference.

III. EVALUATION

A. *Analytical models in Saturation*

The known Bianchi's model [7] allows an accurate evaluation of the saturation throughput of IEEE 802.11 DCF networks under the assumption of ideal channel conditions and considering unlimited retransmissions, by employing a Markov chain. It concludes with the following expression for the saturation throughput:

$$S_t = \frac{P_{tr} P_S E_P}{E_S} \tag{7}$$

where S_t is the saturation throughput defined as the fraction of time the channel is used to successfully transmit payload bits, E_s is the average length of a renewal interval, defined as the time between two consecutive transmissions or the time between two consecutive backoff decrements. E_p is the payload average length, P_{tr} is the probability that at least one user station transmits in a randomly chosen slot time and P_s is the probability of a successful transmission. The detailed derivation of previous parameters can be found in [7].

As mentioned in section II, in (1) and (2) we are introducing errors since we consider that the backoff slots are always shared among the stations. This is true under saturation conditions, that

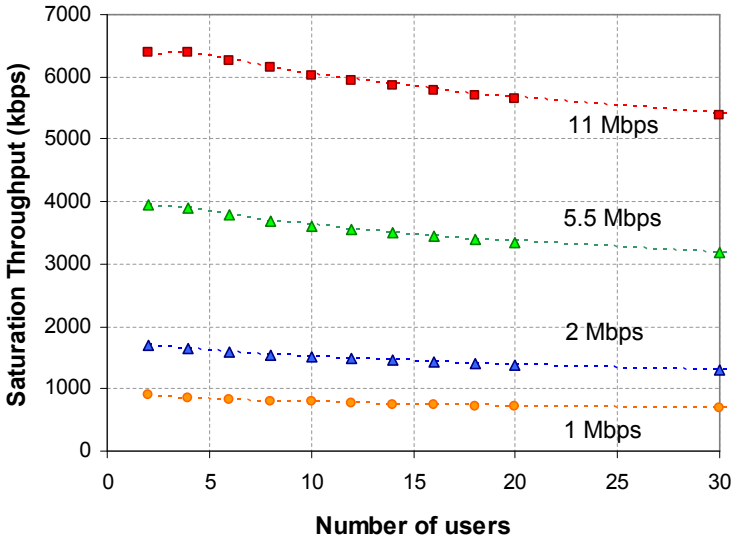

Fig. 1. Saturation throughput for different number of users in a cell (802.11b, MSDU 1500 Bytes)

is, when all stations have always one or more frames ready for transmission. In this situation, all stations decrement their backoff timer at the same time and the one whose timer expires first gains the access; the rest of the stations will decrement the remaining backoff time during the next idle period. Figure 1 shows the results provided by our algorithm (dotted lines) and results obtained with Bianchi's model (symbols). Note that Bianchi's model imposes the following constraints for the scenario: no packet errors, no hidden nodes, and all stations use the same modulation. Results are obtained for the four modulations used in IEEE 802.11b. The MSDU is 1500 Bytes. The average relative error is below 1%.

Chatzmisios *et. al.* introduced the effects of packet errors within Bianchi's model [8]. In fig. 2 we show the results provided by our algorithm (dotted lines) and results obtained with Chatzimisios' model (symbols) for different Packet Error Rates (*PER*).

The graph is obtained with 10 stations and frame size of 1500 Bytes. Recall that P_i is the probability that a frame is not successfully transmitted due to either errors or collisions. This way, the relationship between P_i - equations (4), (6) - and *PER* is:

$$P_i = PER + P_c - PER \cdot P_c \qquad (8)$$

where P_c is the collision probability. In this case, the relative error is also below 1%.

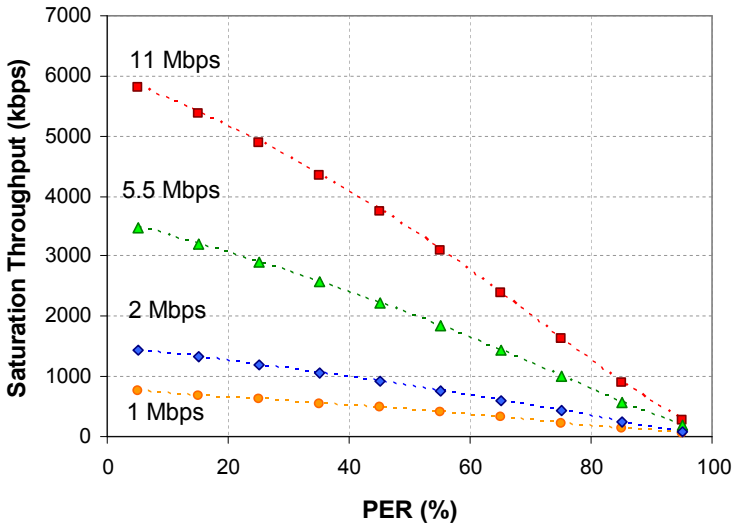

Fig. 2. Saturation throughput for 10 stations and varying PER in a IEEE 802.11b cell, MSDU 1500 Bytes

B. Simulations

As shown in the previous sub-section, backoff slots are shared among saturated stations, but in a more general case, the presence of unsaturated stations avoids such a "synchronization". Two stations will share their backoff slots (or a part of them) only if a packet arrives at both transmission queues at the same time, or during a third station's transmission, or during each other's backoff period (given that the remaining time is greater than DIFS). The probability of this event is low for low offered loads and it is increased as we approach saturation and the number of users is increased. That is to say, the error introduced in our approach is larger with low loaded cells.

The previous evaluation based on analytical models was limited in many ways: it can only be done under saturation conditions, no multi-rate environment and no hidden stations. Conversely, simulations allow a more comprehensive evaluation. We used a simulation tool that closely adheres to all IEEE 802.11 protocol details [9]. The scenario consists of one IEEE 802.11g AP serving various clients. The physical rate used for transmissions depends on the distance between an STA and its selected AP. Figure 3 shows the correlation between rate and distance using the propagation model for semi-open offices proposed in [10].

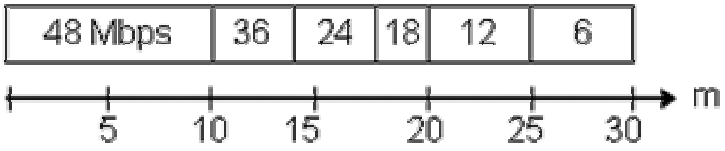

Fig. 3. Coverage radius of an AP for different physical rates

The first set of simulations is intended to demonstrate that the "performance anomaly" of CSMA/CA-based WLANs [5] is effectively captured by our model. The scenario consisted of a single IEEE 802.11g cell with an AP and three stations (*A*, *B* and *C*): *A* used a physical rate of 48Mbps to carry the requirements of a greedy application (FTP), *B* offered traffic described by a 7Mbps UDP CBR source while transmitting at 24Mbps, and *C* increased its traffic demands linearly from 0 to 5Mbps using a physical rate of 12Mbps. Figure 4 shows the carried throughput measured at the application layer for all three stations. Solid lines are drawn for values obtained by simulation and the dotted lines are the values provided by our algorithm. In saturation conditions and according to the algorithm presented in II, the throughput that any of these three stations will obtain (S_j) is about 5Mbps, and this is the value to which each station converges, as seen in fig. 4. As *C* increases its offered traffic, not only it is clear that less capacity is available for greedy stations, but also since *C* is the slowest station, the global throughput is decreased as well.

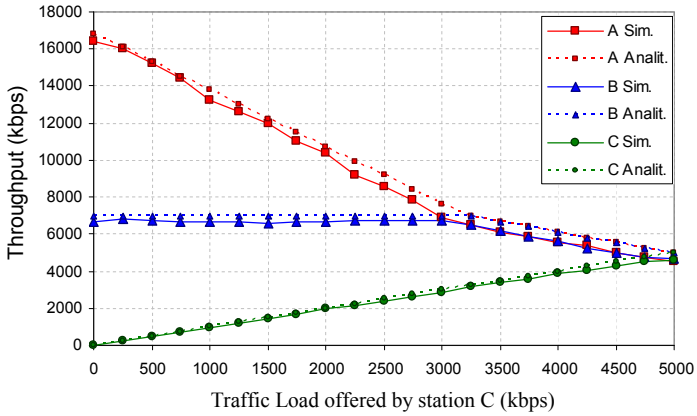

Fig. 4. Throughput of *A* (48M), *B* (24M) and *C* (12M) for different traffic demands in *C*: *A* in saturation and 7Mbps CBR in *B*.

However, the carried traffic can be obtained from direct measurements at the AP. The contribution of the proposed algorithm lies in its capacity to predict the potential throughput that a station will get if associated with a new AP. The next set of simulations compares the predicted AAC with the throughput carried by the newly associated station obtained through simulations. The AP is serving four saturated and mobile stations. In the beginning, the four stations are placed together, 1m away from the AP. At time 0.0s *STA4* starts moving away from the AP, *STA3* at 5.0s, *STA2* at 10.0s and *STA1* at 15.0s (see fig. 5). All four STAs move at 1m/s following opposite directions (N, S, E, and W). As they move away from the AP, the SNR is decreased and the physical rate is adapted dynamically according to the values shown in fig. 3. Figure 6 shows the evolution of this rate adaptation through the simulation. The stations stop moving when they reach the border of the AP's coverage (30m).

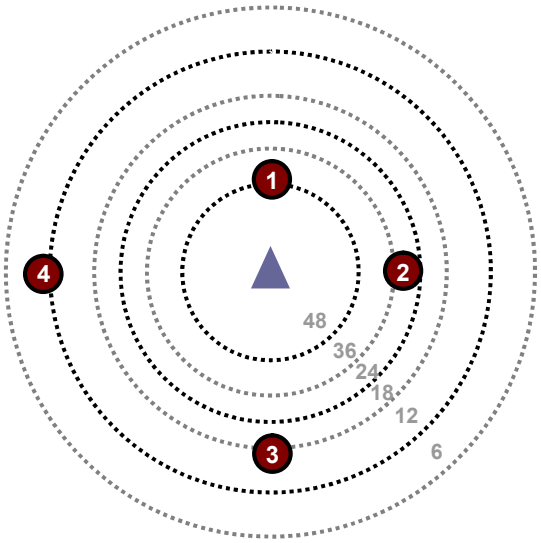

Fig. 5: Picture of the simulated scenario after 35.0s

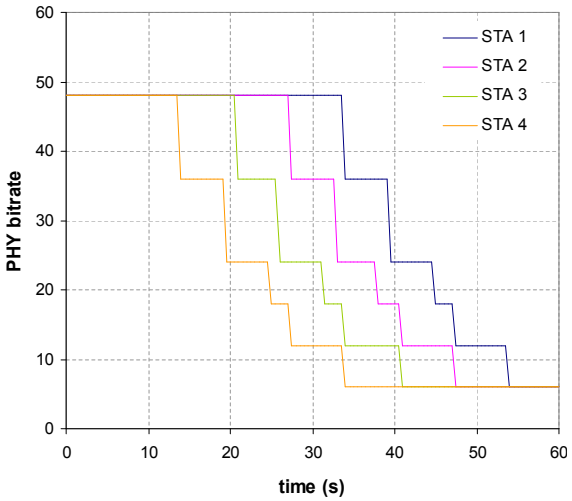

Fig. 6: Effect of rate adaptation on the physical bitrate for four IEE 802.11g STAs moving 1m/s

The algorithm is executed every 0.5s, providing an estimation of AAC. In parallel, a new simulation is run adding a fifth station associated with the AP at the same instants as the AAC estimation is done. Then its saturation throughput is measured twice: one associated at 6Mbps, and the second at 48Mbps. Estimations and simulation results are shown in fig. 7. Logically, the available throughput is decreased by steps every time an associated STA adjusts its physical rate to a slower modulation as shown in fig. 6. The relative error produced by the algorithm that was observed in this scenario is always between 5% and 6%.

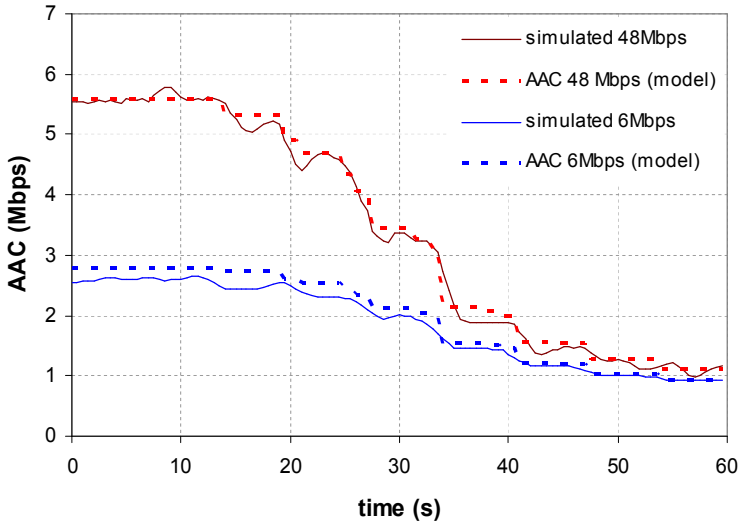

Fig. 7: AAC estimations and simulation results for a fifth 11g associated STA in a scenario with four mobile STAs

The next set of simulations was performed to show the accuracy of the model and the algorithm in a cell with users with varying traffic demands. We define four modes for the stations in a cell:

0) *OFF*: STA is inactive
1) *LOW*: MSDU 500 Bytes and 1 Mbps offered load
2) *MED*: MSDU 1000 Bytes and 2.5 offered load
3) *SAT*: MSDU 1500 Bytes and saturation

In order to see the effects of different modulations, the next scenario consists of 6 stations, each of which is using a different modulation (48, 36, 24, 18, 12 and 6 Mbps); their demands vary according to the behavior shown in fig. 8.

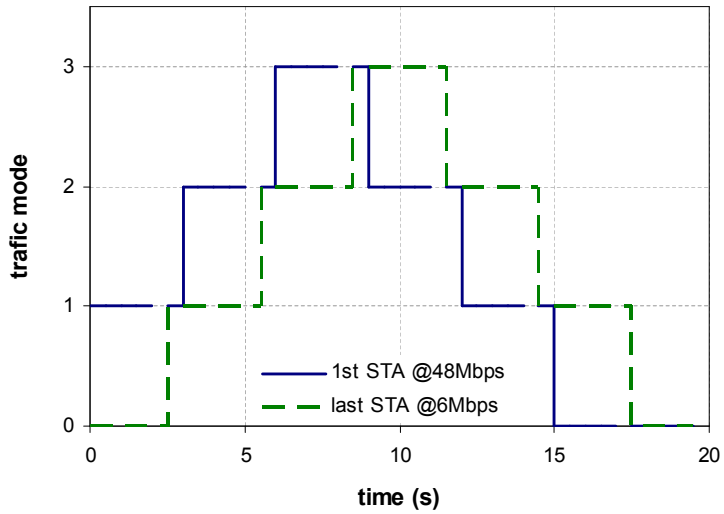

Fig. 8: Variation of the STA's traffic demands (0=OFF; 1=LOW; 2=MED; 3=SAT)

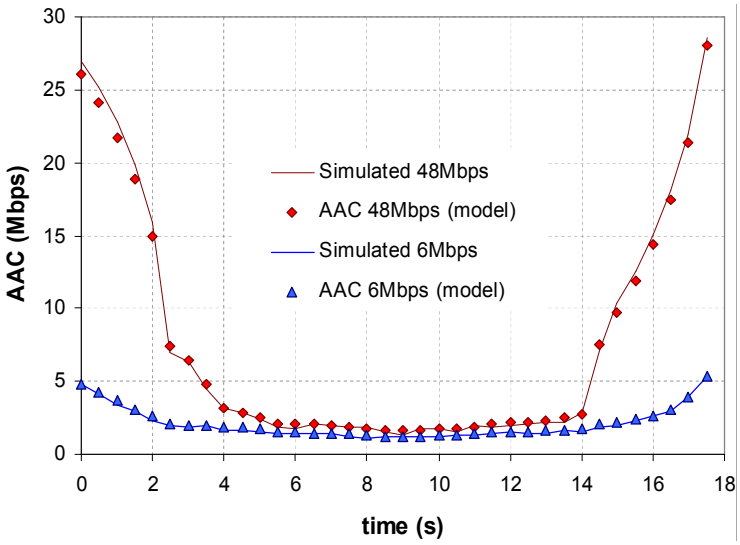

Fig. 9: AAC estimations and simulation results for a seventh 11g associated STA in a scenario with varying traffic demands

STA 1 starts transmitting in mode 1 at time 0.0, STA 2 at 0.5, etc. Figure 9 shows the AAC and simulation measurements for a seventh STA that is associated at 48 Mbps and 6 Mbps. In this case, the relative error is also in the range of 5 – 6%.

Finally a large number of random simulations were performed with increasing number of active users. The stations were placed randomly within the AP's coverage area and their physical rate was chosen according to fig. 3. Their traffic profile was either LOW or MED. Fig. 10 compares the average aggregate throughput obtained through simulations with the values provided by our model. This is the most complete set of simulations since the scenarios evaluated include hidden nodes, channel errors (as a function of each transmission's SINR [10]), different modulations (according to fig. 3) and different traffic demands. Relative error is kept within small values (3 – 5%) for 20 or less stations, but increases to 9% with a very large number of users per cell.

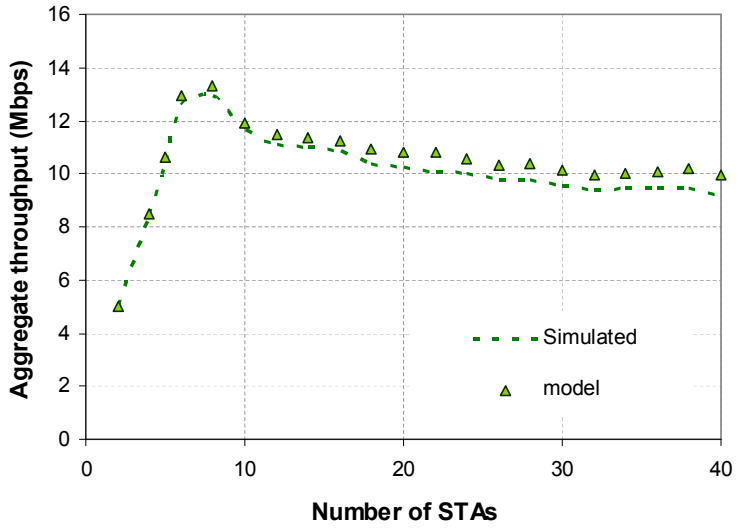

Fig. 10: Aggregate throughput estimations and simulation for randomly generated scenarios with increasing number of STAs

C. *Practical Measurements*

The algorithm is fast and its implementation requires little resources. Therefore it is suitable for running on devices with limited features, such as any commercial AP. In order to prove the viability of our proposal, we implemented the algorithm in a commercial AP running a Linux-based OS: the 4G Systems AccessCube. The algorithm presented in section II was programmed in standard C for Linux and cross-compiled to run on the AccessCube's MIPS architecture. Some of this AP's features include:

- 400MHz MIPS processor
- 32MB flash
- 64MB RAM
- 2 Prism 2.5 based WLAN IEEE 802.11b interfaces
- Linux-based OS: NyLon (kernel $\geq$ 2.4.27)

The AccessCube's OS includes the HostAP driver for the Intersil Prism 2.5 based 802.11b devices on board. This driver provides a helpful collection of statistics which are accessible to applications via the proc filesystem.

Here we present results obtained from measurements in a small testbed. The purpose of these measurements is to validate the algorithm in a real scenario but also to demonstrate that the AAC estimations proposed in section II can be performed in real time by commercial APs.

The testbed topology is shown in fig. 11; there are three stations *A*, *B* and *C*, associated with the same IEEE 802.11b AP. All three stations have different traffic profiles: *A* is the source of an MPEG-2 video stream. The MPEG-2 video codec can be formatted as constant length packets of 188 bytes (this is called Transport Stream) that can build payloads of *k* x 188 bytes. The transport of TS packets over IP/UDP/RTP usually includes 7 TS packets = 1316+40 bytes, in order to approach the Ethernet MTU and maximize efficiency. *A*'s packets are spaced out so that a 2Mbps CBR stream is obtained. In *B*, a greedy application is always trying to send as many 1500 byte UDP packets as possible, while *C* follows a bursty pattern: the average time between consecutive bursts is 20s, the average duration of a burst is 8s; bursts consist of 1000 byte UDP packets in such a way that STA *C* reaches saturation.

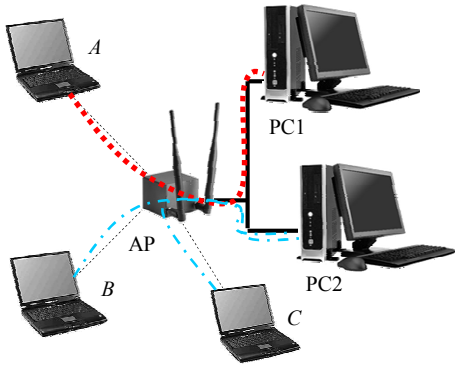

Fig. 11. Testbed for real time: AP runs algorithm and serves three STAs. Connected to PC1 and PC2 with 100M Ethernet

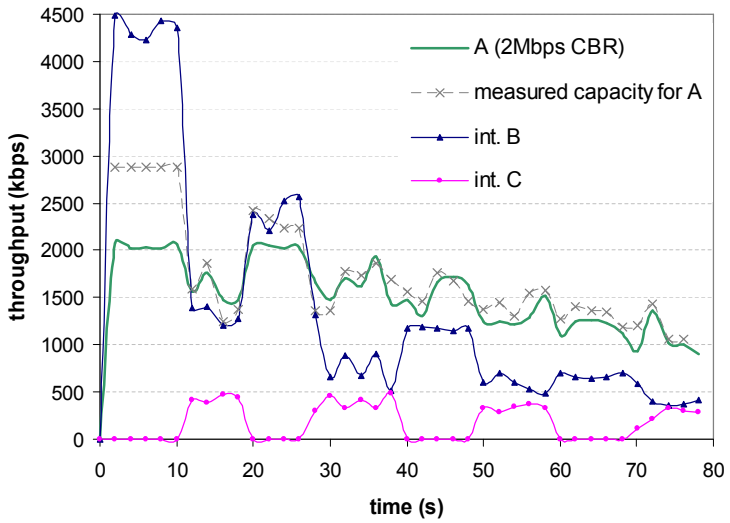

Fig. 12. Real-time capacity measurements for A in the presence of interference from B (greedy) and C (bursty)

The stream originated in *A* is sent to *PC1* while *PC2* is the destination of data flows originated in *B* and *C*. *PC1* and *PC2* are directly connected to the AP by means of a 100Mbps Ethernet segment, so we can guarantee that the bottleneck resides in the air interface. *A* sends frames at 11Mbps, *C* uses 1Mbps and *B* decreases its bitrate one level (modulation) every 20s, starting with 11Mbps. For this experiment, the AP runs the algorithm once per second to provide an estimation of the maximum capacity available for station *A*.

Measurements are shown in fig. 12: during the first 10s station *A* has to compete with only one element, station *B*, which is sending frames at 11Mbps. The measured capacity is greater than *A*'s requirements so *A* can therefore carry all its offered traffic. From 10 to 20s, a burst of *C*'s packets makes the capacity measurements for *A* fall below 2Mbps, which corresponds to the real throughput obtained by *A*'s flow. During the period 20-28s, *B* sends frames at 5.5Mbps, and the capacity for *A* measured at the AP is again greater than *A*'s offered traffic. As it can be seen in the figure, the subsequent measurements are representative of the actual throughput obtained in *A*'s transmissions. This experiment has shown that the implementation in a commercial AP of the algorithm presented in previous sections is able to detect significant capacity fluctuations that affect a given station. Note that the algorithm is slightly overestimating the actual capacity; the values provided are, in average, less than a 10% above the actual throughput measured at the same time instant.

IV. CONCLUSIONS

We have presented a formulation set and an algorithm intended to provide capacity estimations in an IEEE 802.11WLAN cell. The novelty of this approach is that our scheme is able to provide the AAC (Available Admission Capacity) of a cell, that is, the load that a newly associated station could carry if associated to that AP. Moreover, these estimations can be carried out in real time in commercial APs.

Our model is based on several assumptions that may introduce some error in the estimations. The most significant approximation appears when we consider that backoff slots are shared among active stations. While this is true under saturation conditions, the probability of two stations sharing a backoff slot decreases with decreasing load and the number of users. In order to show that the error introduced is small, we performed a comprehensive evaluation of our approach in a wide range of scenarios. The evaluation included analytical and simulation results and also practical measurements.

Our approach was first compared with known analytical models to show that under saturation conditions (no hidden nodes) with and without channel errors, our assumptions were correct (error < 1%). The simulations allowed us the testing of a greater variety of scenarios, including different modulations and traffic demands. The proposed algorithm produced an average relative error near 6%. Finally, the practical measurements are the proof that the algorithm can be run in real time on a commercial AP.

REFERENCES

[1] E. Garcia, D. Viamonte, R. Vidal and J. Paradells, "Achievable bandwidth estimation for stations in multi-rate IEEE 802.11 WLAN cells," in *8th IEEE Symposium on a World of Wireless, Mobile and Multimedia Network*s, WoWMoM'07, June 2007.

[2] A.J. Nicholson,Y. Chawathe, M.Y. Chen, B.D. Noble and D. Wetheral, "Improved Access Point Selection," in *Proc. of MobiSys 2006*. June, 2006.

[3] N. Cranley and M. Davis, "The Effects of Background Traffic on the end-to-end Delay for Video Streaming Applications over IEEE 802.11b WLAN Networks," in *17th IEEE Personal, Indoor Mobile Radio Communications PIMRC*, Sept. 2006.

[4] M. Narbutt and M. Davis, "Gauging VoIP call quality from 802.11 WLAN resource usage," in *Proc. of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM'06*, June, 2006.

[5] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in Proceedings of the 22nd IEEE Annual Conference INFOCOM'03, vol. 2, pp. 836-843, March 2003.

[6] D. Qiao, S. Choi, and K. Shin, "Goodput analysis and link adaptationfor IEEE 802.11a wireless LANs," *IEEE transactions on Mobile Computing*, vol. 1, pp. 278-292, December 2002.

[7] G. Bianchi, "Performance analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535-547, March 2000.

[8] P. Chatzimisios, A. Boucouvalas, and V. Vistas, "Infuence of channel BER on IEEE 802.11 DCF," *Electronics Letters*, vol. 39, pp. 1687-1689, November 2003.

[9] E. Lopez, J. Casademont, J. Cotrina, "Outdoor IEEE 802.11g Cellular Network performance", in Proc. of IEEE Globecom04, November 2004.

[10] T. S. Rappaport, Wireless Communications Principles and Practices. Prentice Hall PTR, second ed., December 2001.